

Geometría Computacional y Bases de Datos

Maria Gisela Dorzán, Susana Esquivel, Edilma Olinda Gagliardi,

Pablo Palmero y Maria Teresa Taranilla

Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis, Argentina
{mgdorzan, esquivel, oli, prpalmero, tarani}@unsl.edu.ar

Gregorio Hernández Peñalver

Departamento de Matemática Aplicada
Facultad de Informática
Universidad Politécnica de Madrid, España
gregorio@fi.upm.es

Resumen

La línea de investigación “Geometría Computacional y Bases de Datos” enmarcada en el proyecto “Tecnologías Avanzadas de Bases de Datos” vincula el estudio de las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas, utilizando los métodos y las herramientas provistas para la resolución de problemas orientados a optimización.

Palabras clave: Bases de Datos, Geometría Computacional, Metaheurísticas, Bases de Datos Espaciales y Espacio-Temporales.

Contexto

El proyecto “Tecnologías Avanzadas de Bases de Datos” de la Universidad Nacional de San Luis tiene como objetivo el estudio de bases de datos avanzadas, y en este marco se investiga el diseño y desarrollo de herramientas que permiten administrar de manera eficiente sistemas de bases de datos no estructuradas.

En el proyecto coexisten tres líneas de investigación, orientadas al desarrollo de nuevos modelos para administrar y recuperar información almacenada en repositorios de datos no estructurados, donde los escenarios de exploración requieren modelos tales como las bases de datos espaciales, bases de datos de

texto, bases de datos de imágenes, bases de datos de sonidos, espacios métricos, entre otros.

En este sentido, en el proyecto “Tecnologías Avanzadas de Bases de Datos” se desarrollan actividades vinculadas al tratamiento de objetos de diversos tipos, estructurados y no estructurados, útiles en diversos campos de aplicación, por ejemplo, robótica, computación gráfica, visión artificial, sistemas de información geográfica, diseño asistido por computadora, computación móvil, motores de búsqueda en internet, entre otras, y que se relacionan en tales bases de datos.

Así, surge el estudio de modelos como las bases de datos espaciales y bases de datos espacio-temporales y la construcción y manipulación de diferentes objetos y estructuras geométricas de utilidad en diversas áreas de aplicación. En particular, las estructuras geométricas que se estudian deben cumplir con propiedades deseables y algunos de los problemas relacionados con la optimización de las mismas son problemas NP-duros, por tanto en la búsqueda de soluciones aproximadas se recurre a técnicas metaheurísticas.

En la línea de investigación “Geometría Computacional y Bases de Datos” se vinculan temáticas que surgen de las disciplinas Bases

de Datos, Geometría Computacional y Metaheurísticas.

El trabajo de investigación se desarrolla en forma conjunta con investigadores afines de proyectos locales y de universidades extranjeras (Universidad Politécnica de Madrid - España, Universidad Veracruzana - México, entre otras) mediante convenios de cooperación institucional. En este trabajo, se presentan temas de interés junto con las propuestas más recientes de trabajo.

Introducción

La Geometría Computacional estudia el diseño de algoritmos para resolver problemas geométricos. En esta disciplina se identifican conceptos, propiedades y técnicas que apuntan al desarrollo de algoritmos eficientes. Esto conduce al análisis y estudio de estructuras de datos geométricas y problemas relacionados con dichas estructuras [BCKO08]. Los problemas se estudian desde un punto de vista combinatorio o algorítmico. Los aspectos algorítmicos de los problemas aparecen en la búsqueda de soluciones, exactas o aproximadas, para una estructura geométrica particular. En este contexto aparecen problemas que pertenecen a la clase de complejidad NP, o bien, son problemas para los cuales no se conocen algoritmos eficientes que permitan alcanzar soluciones exactas. En determinadas circunstancias interesa obtener soluciones a tales problemas, aunque sean aproximadas.

Algunos de los problemas refieren a ciertas configuraciones geométricas obtenidas a partir de un conjunto de puntos u objetos en el plano y se busca optimizar propiedades que miden la calidad de dichas configuraciones. Los objetivos están orientados básicamente a construir estructuras geométricas que satisfagan ciertas propiedades vinculadas a problemas de optimización para los cuales hay estudios acerca de su naturaleza NP-dura. Para ello, se propone la aplicación de técnicas metaheurísticas. Una metaheurística es una metodología de alto nivel que puede ser utilizada como una estrategia que guía las heurísticas subyacentes para resolver

problemas de optimización específicos, combinando inteligentemente diferentes conceptos de diversos campos. Estos métodos son simples de implementar y han demostrado poder encontrar, de forma eficiente, buenas soluciones para problemas de optimización NP-duros [MF04] [BFM97].

Los problemas en los que se trabajaron corresponden a estructuras geométricas, obtenidas a partir de un conjunto de puntos u objetos en el plano, para las cuales se busca optimizar alguna propiedad que permita valorar la calidad de dichas estructuras. Los criterios de optimización consisten en minimizar o maximizar alguna propiedad de las estructuras geométricas.

Entre los problemas de optimización estudiados destacamos: la Triangulación de Peso Mínimo (*Minimum Weight Triangulation, MWT*) y la Pseudo-Triangulación de Peso Mínimo (*Minimum Weight Pseudo-Triangulation, MWPT*). Dichos problemas minimizan la suma de las longitudes de las aristas proveyendo así una medida de calidad de la estructura. La complejidad del cálculo de MWT fue uno de los problemas abiertos más estudiados en Geometría Computacional hasta que se demostró que la construcción de MWT es un problema NP-duro [MR06]. Con respecto a la complejidad de MWPT aún no está resuelta. Existe una 12-aproximación de una pseudo-triangulación que puede ser calculada con complejidad de $O(n^3)$ y una aproximación de $O(\log n \cdot w(\text{MST}))$ de una pseudo-triangulación de peso mínimo en un tiempo de $O(n \log n)$ donde $w(\text{MST})$ es el peso del Árbol de Expansión Mínimo (*Minimum Spanning Tree, MST*) el cual es un subconjunto de la estructura obtenida. [GL07].

Otro problema en estudio es la Triangulación de Dilación Mínima (*Minimum Dilation Triangulation, MDT*) donde la dilación mide la calidad de conexión entre puntos de la triangulación. La dilación entre un par de puntos, u y v , de una triangulación T se define como la razón entre el camino más corto entre u y v en T y la distancia Euclídea entre dicho par de puntos. La máxima dilación entre todos los pares de puntos en T se llama la dilación de

T ($\Delta(T)$). La mejor posible dilación de cualquier triangulación de un conjunto de puntos S se conoce como dilación de S ($\Delta(S)$). La complejidad del problema de encontrar $\Delta(S)$ para un conjunto de puntos S se desconoce y por lo tanto, el trabajo se centra en el desarrollo de algoritmos aproximados para encontrar triangulaciones de buena calidad de dilación mínima. Los problemas de optimización relacionados a triangulaciones y pseudo-triangulaciones son de interés debido a su utilidad en diversos campos: detección de colisiones, visibilidad y vigilancia, problemas de rigidez de estructuras, entre otros.

La complejidad de los problemas antes mencionados justifica la utilización de técnicas metaheurísticas que permitan obtener soluciones aproximadas a las óptimas.

Línea de investigación, Desarrollo e Innovación

En la línea de investigación se propone el estudio de optimizaciones de estructuras geométricas relacionadas con las bases de datos mencionadas, el desarrollo de herramientas para la visualización de estructuras geométricas y aplicaciones vinculadas a las bases de datos espacio-temporales. En particular, los problemas de interés en la línea de investigación son aquellos vinculados al diseño de índices espacio-temporales para resolver en forma integral consultas espacio-temporales y su vinculación con la problemática de objetos móviles; y al tratamiento de problemas geométricos NP-duros. Para estos últimos, en la búsqueda de soluciones se utilizan las técnicas metaheurísticas que proporcionan estrategias adecuadas para la obtención de soluciones aproximadas de múltiples problemas de optimización.

Como objetivos específicos de estudio podemos enumerar los siguientes:

- Estudio de configuraciones geométricas generales de puntos en el plano, que cumplan medidas de calidad como: peso, dilación, área, entre otras. Optimización de triangulaciones y pseudo-triangulaciones, considerando medidas

de calidad mínimas o máximas, aplicando técnicas metaheurísticas y otras estrategias algorítmicas.

- Diseño y desarrollo de herramientas para la generación, visualización y manipulación de diferentes configuraciones geométricas de conjuntos de puntos en el plano.

- Desarrollo de aplicaciones utilizando herramientas de Geometría Computacional y Bases de Datos Espacio-Temporales.

Resultados y Objetivos

Para la resolución de problemas de optimización de triangulaciones y pseudo-triangulaciones de peso mínimo se aplicaron las técnicas metaheurísticas: Optimización basada en Colonia de Hormigas (*Ant Colony Optimization*, *ACO*) y Recocido Simulado (*Simulated Annealing*, *SA*), técnicas determinísticas Voraces (*Greedy*) y Triangulación de Delaunay. Se llevó a cabo el estudio, adecuación y evaluación experimental de las técnicas metaheurísticas mencionadas para la búsqueda de triangulaciones y pseudo-triangulaciones que cumplan la propiedad. Se diseñaron generadores de instancias de problema para ser utilizados en la evaluación experimental. Dichas colecciones se encuentran disponibles en el sitio de la línea de investigación (www.dirinfo.unsl.edu.ar/bd2/GeometriaComp/).

Los resultados obtenidos para los problemas MWT y MWPT utilizando la técnica metaheurística ACO fueron publicados en [DGLH11b] [GDLH11] [DGLH12]. Se llevó a cabo una evaluación experimental y análisis de los resultados obtenidos con la técnica Recocido Simulado [DGLH11a] [DGLH11c].

Para el problema MDT todavía no se conoce un algoritmo que lo resuelva en tiempo polinomial y tampoco se ha demostrado que sea NP-duro. Las técnicas aplicadas para este problema fueron: Greedy, Local Search, Iterated Local Search, Simulated Annealing y Random Local Search. Para cada estrategia se propuso un conjunto de operadores adecuados. Debido a la complejidad que implica la puesta a punto de los parámetros de técnicas metaheurísticas se utilizó Optimización de

Parámetros Secuencial (*Sequential Parameter Optimization, SPO*) para el ajuste de los parámetros requeridos por Simulated Annealing. Se realizó un análisis experimental en el cual se compararon dichos algoritmos con otras técnicas, como por ejemplo, Delaunay. Se crearon las instancias de prueba, ya que para estos problemas no se encontraron disponibles ningún tipo benchmark con el cual comparar nuestros resultados. Las conclusiones fueron afirmadas desarrollando un estudio estadístico aplicando diferentes test estadísticos y métodos de visualización [DLMH14].

Los resultados obtenidos en el tratamiento de estos problemas se plasmaron en dos tesis de doctorado.

Por otra parte, se desarrolló una herramienta para la generación y visualización de triangulaciones, pseudo-triangulaciones y poligonizaciones de conjuntos de puntos en el plano. Además, se finalizó con la implementación de una aplicación en el ámbito de la Salud para el seguimiento de focos epidémicos utilizando base de datos espacios-temporales y herramientas de Geometría Computacional. El desarrollo de las herramientas mencionadas se plasmó en dos trabajos finales de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas [PDG13], [GPDGT14].

Como trabajo futuro, considerando los problemas sobre configuraciones geométricas en estudio, se pretende continuar con el estudio de los problemas de optimización considerando otras técnicas metaheurísticas, adecuadas para su resolución.

Formación de Recursos Humanos

La formación del grupo de trabajo en la Universidad Nacional de San Luis, se consolida con actividades de cooperación mutua e intercambio recíproco de información científica, tecnología y desarrollo de nuevos conocimientos con investigadores locales y de otras universidades.

Entre las actividades más destacadas se pueden mencionar:

- Formación de recursos humanos plasmada dos Tesis Doctorales (UNSL), tres Licenciados en Ciencias de la Computación (UNSL).
- Obtención de becas de investigación tipo I y tipo II de CONICET.
- Realización y dirección de pasantías de investigación con docentes de otras universidades.
- Actividades de formación académica, dictado de cursos de posgrado y de especialización.
- Actividades de divulgación científica, conferencias, publicaciones en congresos y revistas en el ámbito nacional e internacional.

Por lo expuesto anteriormente, la línea de investigación “Geometría Computacional y Bases de Datos” tiene como propósito continuar con el trabajo de investigación para consolidar y alimentar líneas de estudio en la disciplina que permitan fortalecer y afianzar la formación de recursos humanos locales, proponiendo actividades de formación académica, de investigación, desarrollo, y otras actividades académico-científicas vinculantes para que ello represente un aporte en la comunidad científica.

Referencias

- [BFM97] Bäck T., Fogel D., Michalewicz Z. *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press. 1997.
- [BCKO08] de Berg, M., Cheong O., van Kreveld, M., Overmars, M., *Computational Geometry: Algorithms and Applications*. 3rd edition, Springer-Verlag, Heidelberg, 2008.
- [DGLH11a] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approaches for MWT and MWPT Problems*. XVII Congreso Argentino de Ciencias de la Computación 2011 (CACIC 2011), 2011
- [DGLH11b] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Using ACO metaheuristic for MWT problem*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011

- [DGLH11c] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Metaheuristic approaches for MWT and MWPT Problems*. XIV Encuentros de Geometría Computacional. Páginas: 79-82. 2011
- [DGLH12] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approximations on Minimum Weight Triangulations and Minimum Weight Pseudo-Triangulations using Ant Colony Optimization Metaheuristic*. Fundamenta Informaticae. ISSN: 0169-2968 (Print), 1875-8681 (Online). Volume 119, number 1, pp 1-27.
- [DLMH14] M. G. Dorzán, M. G. Leguizamón, Efrén Mezura-Montes, G. Hernández Peñalver *Approximated algorithms for the Minimum Dilation Triangulation Problem*. Journal of Heuristics. DOI 10.1007/s10732-014-9237-2. Print ISSN 1381-1231. Online ISSN 1572-9397. Publisher Springer US. 2014.
- [GDLH11] Gagliardi E. O., Dorzán M. G., Leguizamón M. G. y Hernández Peñalver. G.; *Approximations on Minimum Weight Pseudo-Triangulation problem using Ant Colony Optimization*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011
- [GL07] Gudmundsson J., Levkopoulos C.; *Minimum weight pseudo-triangulations*. Computational Geometry. Theory and applications. Elsevier Vol. 38- pages 139-153, 2007.
- [GPDGT14] Guasch, M.M; Piergallini, M.R ; Dorzán, M.G.; Gagliardi, E.O.; Taranilla, M.T.; *“Una herramienta para el análisis y seguimiento de focos epidémicos”* en Anales del 17° Concurso de Trabajos Estudiantiles en 43 Jornadas Argentinas de Informática. Páginas: 35-45- Universidad de Palermo, Buenos Aires, Argentina, 2014.
- [MR06] Mulzer W., Rote G. *Minimum weight triangulation is NP-hard*. In Proceedings of the 22nd Annual ACM Symposium on Computational Geometry. 2006.
- [MF04] Michalewicz Z., Fogel D., *How to Solve It: Modern Heuristics*, Springer, 2004.
- [PDG13] Palmero, P.R., Dorzán, M. G., Gagliardi E.O., *Una Herramienta para la Manipulación de Configuraciones Geométricas*, 42° Jornadas Argentinas de

Informática e Investigación Operativa - 16° Concurso de trabajos de fin de carrera. Universidad Nacional de Córdoba, Córdoba, Argentina. 2013. ISSN: 1850-2946.